



ALMA MATER STUDIORUM
UNIVERSITA' DI BOLOGNA

De novo genome assembly and functional genomics of the *Pelobatrachus nasutus*

as a phenotypic plasticity, and
antimicrobial defenses model

Marco Cuscunà



Background knowledge



Pelobatrachus Nasutus in his habitat [1].

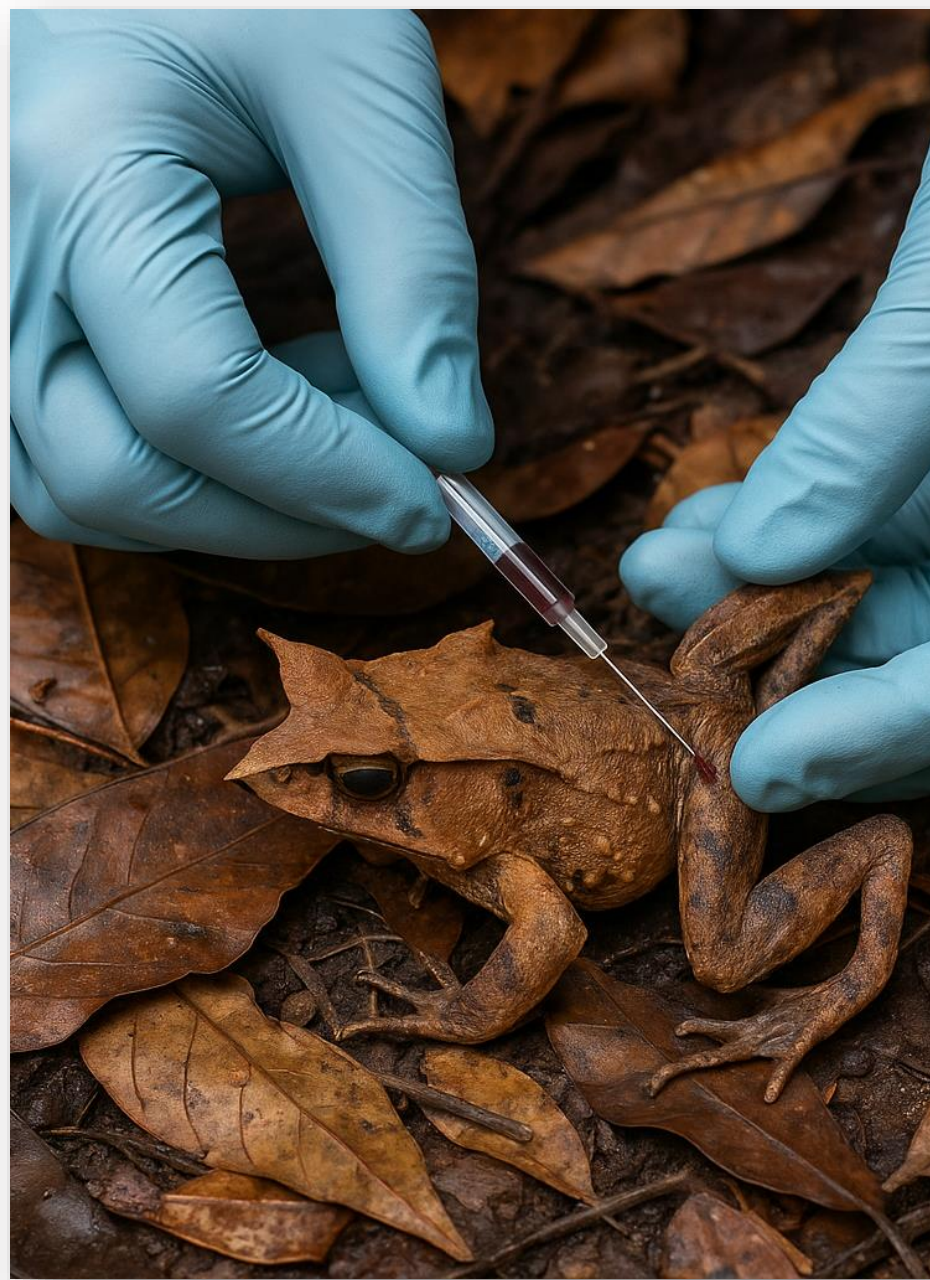
It is a Southeast Asian amphibian known for its striking leaf like camouflage, with the eyelid projection and a characteristic textured skin.

This species exhibits notable **phenotypic plasticity** and produces **Antimicrobial Peptides (AMPs)** interesting for biomedical purposes.



What is **the aim of the study?**

Generate the first and annotated reference genome of Pelobatrachus nasutus, highlighting potential SNPs involved in differences between different habitats (focusing on genes involved in morphological variations, and AMPs production).



Representative image of *Pelobatrachus nasutus*' blood sample extraction, for illustrative purposes only [2].

Sampling sites in Sarawak, Malaysian Borneo: Gunung Gading National Park (protected forest) and Wilmar Oil Palm Plantation (fragmented habitat) [3].



Sample Collection

Reference Genome samples

Samples were collected from blood to obtain High Molecular Weight (HMW) DNA, and from tissue biopsies to perform RNA-seq for gene annotation.

Population genomics

Samples were collected from 10 individuals: 5 from Gunung Gading National Park and 5 from the WILMAR Oil Palm Plantation. RNA was extracted to enable population-level differential analysis.

2. Image source: <https://www.canva.com/ai-image-generator/>

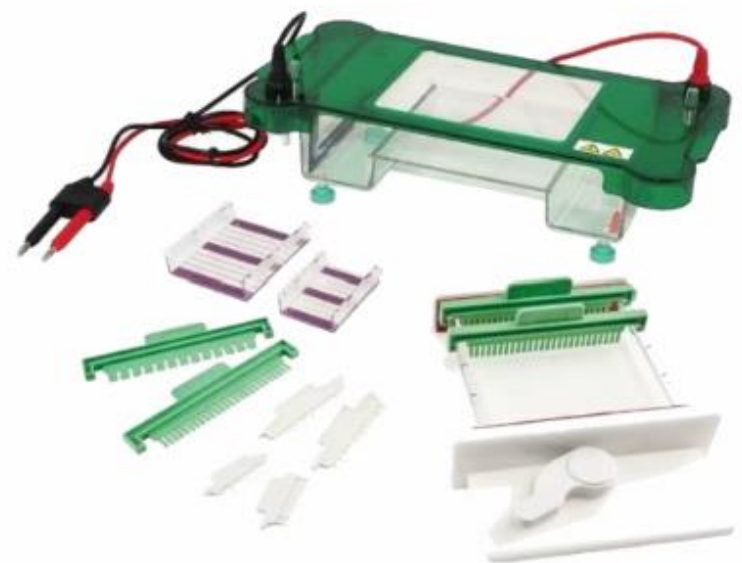
3. Edited from: https://commons.wikimedia.org/wiki/File:Relief_Map_of_Borneo.png



RNA/DNA extraction

Qiagen Genomic-tip was used to extract DNA, while Qiagen Rneasy kit to perform RNA extraction.

Qubit fluorometer, was used to make accurate quantification of double-stranded DNA, and spectrophotometer with electrophoresis to evaluate purity and sizes of the fragments.



Reference Genome Sequencing

Sequencing was performed with PacBio HiFi platform which provided highly accurate reads (with a quality value >30 and coverage of 30–40x).

PacBio



Representative image of PacBio Sequel IIe Hi-Fi System, for illustrative purposes only [4].

4. Image source:
<https://www.globenewswire.com/NewsRoom/AttachmentNg/ec7721b6-506f-416d-a80b-2837542bdb85/en>

Contig Assembly and scaffolding



Representative image of Illumina NovaSeq PE150 System, for illustrative purposes only [5].



hifiasm

To assemble the reads was used hifiasm an assembler specifically optimized for PacBio HiFi data.



Illumina NovaSeq sequencing

To achieve chromosome level genome, Hi-C libraries were prepared from fresh tissue DNA sequenced with Illumina system.



3D-DNA and Juicebox

Hi-C contact maps libraries are used in combination with 3D-DNA and Juicebox for scaffolding.



Assembly quality

Assessed using BUSCO, QV and N50.



RepeatModeler2 and RepeatMasker

RepeatModeler2 is used to identify de novo repeated elements and RepeatMasker to mask them preventing spurious gene predictions.



RNA-sequencing and STAR

Was performed RNA-seq on tissue skin samples of the reference individual, and reads were mapped through STAR against the assembled genome.



BRAKER2

Gene prediction was done following BRAKER2 pipeline, combining RNA-seq evidences with AUGUSTUS ab initio models.



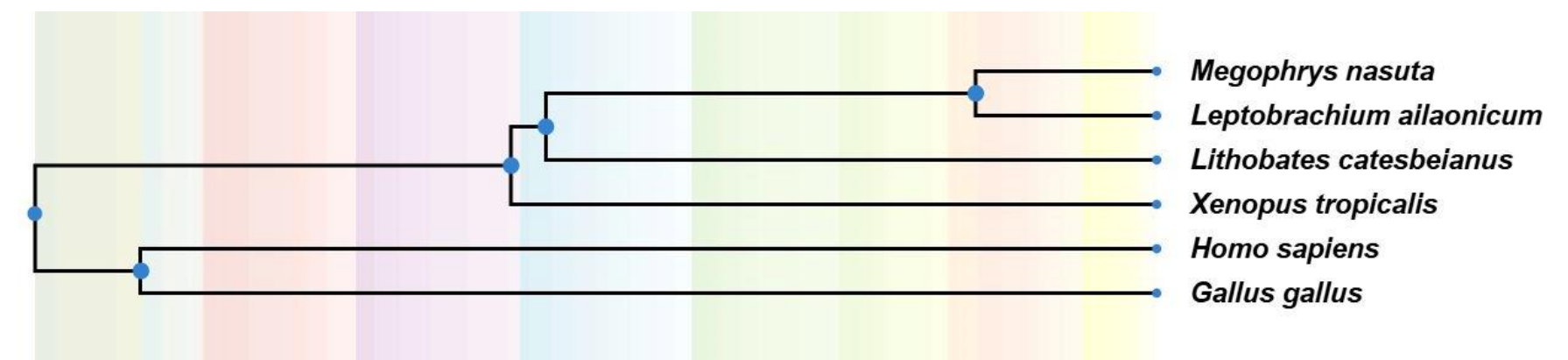
MAKER

To refine it, was used MAKER pipeline to integrate these results with homology from related *Leptobrachium Ailaonicum* genome already studied.

Functional annotation



Leptobrachium Ailaonicum in his habitat [6]



TimeTree-based phylogeny showing the evolutionary placement of *Pelobatrachus* (*Megophrys*) *nasutus* relative to the closely related *Leptobrachium ailaonicum*

Population analysis



Comparison of esemplars of Pelobatrachus Nasutus, one coming from protected environment (on the right) and the other one from fragmented environment (on the left)[7].



BWA-MEM2

Reads from the genome of the 10 individuals coming from the two habitats, were sequenced with Illumina NovaSeq and mapped against the reference genome through BWA-MEM2.



GATK (Genome Analysis Toolkit)

Variant calling was performed GATK that reconstructed the local haplotypes, output SNP and indel calls.



Address medium coverage genotypes

To address this population genetic analysis was based on genotype likelihoods rather than hard genotype calls.



PCA, Fst and BayPass

PCA and Fst are used to assess the presence of a significant differences between the two population. While BayPass detected correlations between allele frequencies and ecological variables.

Expected Results And Data Submission

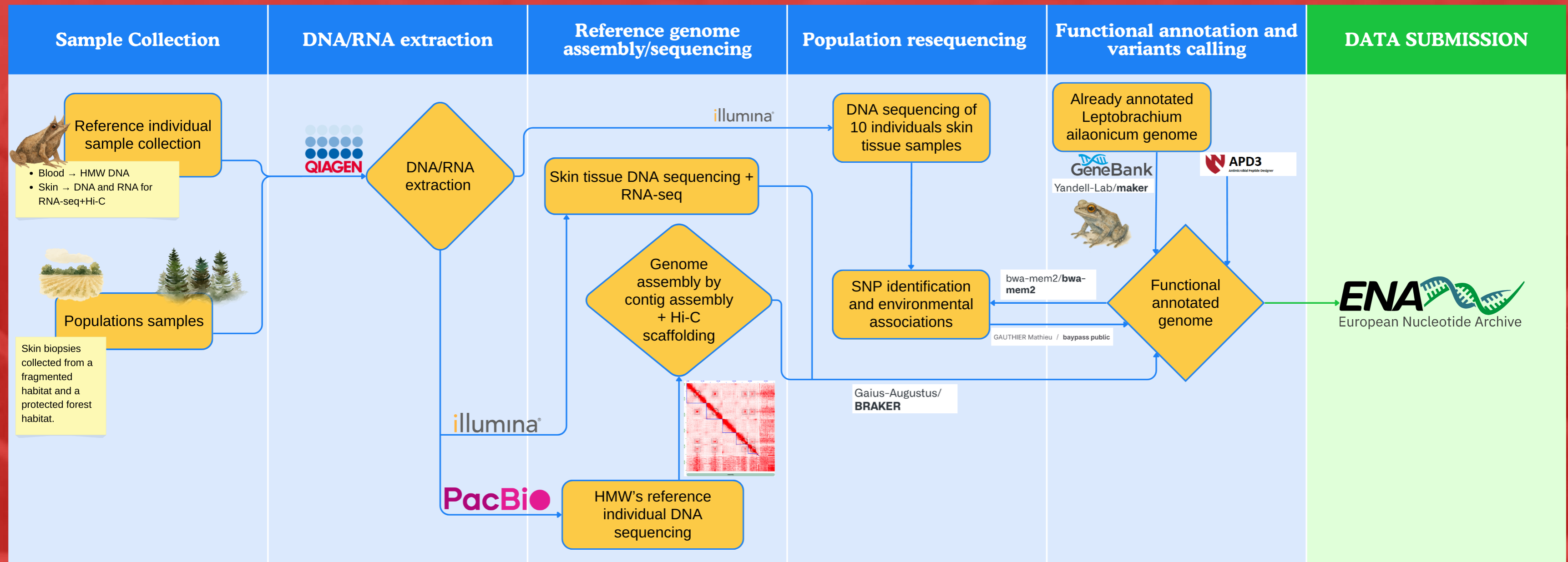
- ✓ **High-quality and annotated reference genome**
- ✓ **Detection of genes potentially related to skin adaptation highlighting the phenotypic plasticity**
- ✓ **Annotation of genes potentially involved in production of AMPs for biomedical purposes**

All data at the end of the experiment will be submitted in **ENA (European nucleotide archive)** to ensure reproducibility and reuse by the scientific community.



Pipeline Diagram

De novo genome assembly and functional genomics of the *Pelobatrachus nasutus* as a phenotypic plasticity, and antimicrobial defenses model



Estimated costs

Budget: **\$100.000**

ITEM	Estimated Cost	Notes
Fieldwork & permits	4800	Sampling in Gunung Gading NP & Wilmar plantation; ABS/Nagoya compliance
DNA extraction (HMW + resequencing samples)	3250	Qiagen Genomic-tip+consumables
RNA extraction (skin only)	1420	RNAlater + Qiagen RNeasy kits
PacBio HiFi sequencing	34780	30-40x coverage, 1 HQ individual
Hi-C sequencing	8230	Hi-C libraries from fresh tissue, sequenced on illumina NovaSeq PE150
Illumina resequencing (10 individuals, 15x)	14650	Illumina NovaSeq PE150
RNA-seq (skin sample, 2-3 replicates)	6180	Illumina NovaSeq PE150
Computational costs & Cloud storage	11740	Assembly, annotation, population genomics analyses
Consumables & contingency	8560	Tubes, barcoding, dry shipper
TOTAL	93610	Inside the 100000 \$ budget.